THE STAGES OF EARLY ARITHMETIC LEARNING: A CONTEXT FOR LEARNING TO PROFESSIONALLY NOTICE

Molly H. Fisher
University of Kentucky
molly.fisher@uky.edu

Edna O. Schack
Morehead State University
e.schack@morehead-st.edu

Jonathan Thomas
Kentucky Center for Mathematics
thomasj13@nku.edu

Sara Eisenhardt
Northern Kentucky University
eisenhards1@nku.edu

Margaret Yoder
Eastern Kentucky University
margaret.yoder@eku.edu

Janet Tassell
Western Kentucky University
janet.tassell@wku.edu

The goal of this study is to develop the professional noticing abilities of preservice elementary teachers in the context of the Stages of Early Arithmetic Learning. In their mathematics methods course, the preservice elementary teachers participated in a researcher-developed multi-session module that progressively nests the three interrelated components of professional noticing—attending, interpreting, and deciding. A pre- and post-assessment was administered to measure their change in the three components of professional noticing. The preservice elementary teachers demonstrated significant growth in all three components.

Keywords: Elementary School Education; Learning Trajectories; Number Concepts and Operations; Teacher Education–Preservice

Introduction and Literature

This study examines the extent to which an innovative learning experience focused on the professional noticing of children's early numeracy develops Preservice Elementary Teachers' (PSETs) capacity to attend to, interpret, and make effective instructional decisions related to the mathematical thinking of children. The study relies upon a module, *Noticing Numeracy Now (N^3)*, developed by the researchers and based on literature in the areas of professional noticing (Jacobs, Lamb, & Philipp, 2010) and the Stages of Early Arithmetic Learning (SEAL) (Steffe, von Glasersfeld, Richards, & Cobb, 1983; Steffe, Cobb, & von Glasersfeld, 1988; Steffe, 1992). Specifically, we intend to investigate the following research question: To what extent can teacher educators facilitate the development of PSETs' capacity to professionally notice children's mathematical thinking?

Professional Noticing

Professional noticing is an ability to recognize and act on key indicators significant to one's profession. The literature exploring the impact of professional noticing in mathematics teaching grew significantly in recent years. Sherin and van Es (2009) examined teacher video clubs as a tool for analyzing the club participants' classrooms and found that using focused noticing as a lens for learning about teaching was productive beyond the video club, impacting the teachers' instructional practices. Star and Strickland (2008) demonstrated improvements in preservice teachers' ability to attend to the salient features of a secondary mathematics classroom. Numerous professional development modules incorporated the use of video to focus observers' attention on children's mathematical thinking (Carpenter, Fennema, Franke, Levi, & Empson, 1999; Schifter, Bastable, & Russell, 2000; Seago, Mumme, & Branca, 2004). Sherin, Jacobs, and Philipp's recently edited volume (2011) contributed to the compounding evidence of both the need and the value of professional noticing to effective mathematics teaching.

Recent evidence shows that teachers' attention to children's mathematical thinking can positively affect student learning (Carpenter et al, 1999; Kersting et al, 2010); however, such attention is just one component skill of *professional noticing* of children's mathematical thinking as defined by Jacobs, Lamb, and Philipp (2010). They conceptualized professional noticing as "a set of three interrelated skills: *attending* to children's strategies, *interpreting* children's understandings, and *deciding* how to respond on the basis of children's understandings" (p. 172). Their research examined the professional noticing of

Van Zoest, L. R., Lo, J.-J., & Kratky, J. L. (Eds.). (2012). *Proceedings of the 34th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education.* Kalamazoo, MI: Western Michigan University.

preservice teachers as well as three groups of in-service teachers all having 12–14 years of teaching experience but varying degrees of professional development. The results indicate that teaching experience alone does not develop all three components of professional noticing. Teachers with 12–14 years experience, but no sustained professional development, aligned more closely with preservice teachers on this construct, especially on the deciding component. Star and Strickland (2008) further contend that developing the skill of professional noticing must be an early focus of teacher preparation programs given the importance of skillful, nuanced observation in learning to teach.

Practice-based Teacher Preparation

Preparing for the profession of teaching requires opportunities to practice one's teaching. Current trends in teacher preparation focus on practice-based teacher preparation but with varying degrees of guidance on the particulars of what should be practiced and where practice should take place. Grossman, Compton, Igra, Ronfeldt, Shahan, and Williamson (2009) found that teacher preparation practices focused heavily on the preactive aspects of teaching, such as lesson and unit planning, and less on interactive or reflective aspects. The interactive and reflective aspects include the nearly invisible decisions based on professional noticing that must be made 'in the moment' of teaching. We expect novice teachers to observe the professional noticing in real-time classrooms, but without explicit guidance, novice teachers may not observe it. Videos of teaching/learning exchanges are representations of practice, one of three pedagogies of practice proposed by Grossman and her colleagues. Such representations can potentially provide powerful settings for learning the practice of teaching and in doing so can provide a scaffold for subsequent practice in actual classroom-based contexts.

Stages of Early Arithmetic Learning

There is consensus that the term "numeracy" is grounded in an understanding of fundamental and foundational aspects of number and operations (Mulligan, Bobis, & Francis, 1999; National Council of Teachers of Mathematics, 2006) and is a significant content strand for PSET exploration. Steffe and his colleagues have provided a useful model for structuring the mathematics of the N^3 module. Born of longitudinal teaching experiments, the Stages of Early Arithmetic Learning (SEAL) hypothesize a progression for the development of quantitative understanding (Steffe, von Glasersfeld, Richards, & Cobb, 1983; Steffe, Cobb, & von Glasersfeld, 1988; Steffe, 1992; Wright, Martland, & Stafford, 2006). This progression includes the following levels: Emergent, Perceptual, Figurative, Initial Number Sequence, Intermediate Number Sequence, and Facile. Given the supporting methodology, SEAL is exemplary of "learning trajectories built upon natural developmental progressions identified in empirically based models of children's thinking and learning" (Clements, 2007, p. 45).

Methodology

Participants

The participants in this study were preservice elementary teachers (PSETs) enrolled in one of three participating universities; two are regional universities and one is a "Research Very High" university (Carnegie Foundation for the Advancement of Teaching). All universities are public institutions and the participants (n = 94) represent a cross-section of the general population of a state in the east-central United States. Participants were enrolled in elementary mathematics methods courses at their respective universities and the module was a component of this course.

Module Description

The module consists of multiple in-class sessions during which the three components of professional noticing are developed in the context of the Stages of Early Arithmetic Learning. The three components of professional noticing, attending, interpreting, and deciding, are nested through the module (Boerst, Sleep, Ball, & Bass, 2011). The first two sessions focus on the development of attending only. Subsequent sessions further develop attending while progressively layering in interpreting and deciding. Integrated with the nested development of professional noticing, the Stages of Early Arithmetic Learning gradually unfold through video clip representations of practice. The researchers of the N^3 project intentionally chose video around early number sense for two reasons: (1) video is a representation of practice that provides opportunities to explicitly attend to and discuss salient features of children's mathematical thinking that can go unnoticed by novices in a real-time classroom setting, and (2) early number sense is an area of mathematics with which PSETs are generally comfortable so the mathematics itself would not be a barrier to the examination of children's mathematical thinking.

The video clips are diagnostic interviews with children conducted by teacher educators or a former PSET. A significant number of video clips are from one author's dissertation research of quantitative mental imagery (Thomas, 2010). PSETs are asked to respond to the videos in various ways, including writing about what they attend to in the video and engaging in discussion with a partner, with a small group, and in whole class discussions led by the instructor. At the beginning of the module, the discussion prompts are more general, such as asking them to observe the physical actions and verbal exchanges taking place. As the module continues, the PSETs learn to focus on the salient features of students' mathematical actions and words. In addition to these salient features, PSETs' attention is drawn to teacher moves and the mathematics of the tasks. As the sessions progress, PSETs learn to interpret the salient features in terms of SEAL and finally they learn to make decisions about next steps, either diagnostic or instructional. Between the sessions, the PSETs have homework including articles to read and videos to watch. The culminating experience is an assignment that requires the PSETs to conduct at least one diagnostic interview with a child. The interview assignment varies across the universities. Some are assigned immediately following the module, and others are assigned much later in the semester, dependent upon each university's field placement schedules.

Professional Noticing Measures

A pre- and post-assessment designed to measure a PSET's ability to apply professional noticing to a video clip representation of practice was developed. The pre-assessment was administered within a week of the start of the semester at all participating institutions. The professional noticing measure consists of a brief video clip (25 seconds) in which the interviewer poses a partially screened task that goes beyond finger range. The task is a comparison task, where the difference between two sets is unknown (Carpenter et al, 1999). After viewing the video twice, PSETs were asked to respond to three prompts, each related to one of the three aspects of professional noticing—attending, interpreting, and deciding. The three prompts, drawn from the work of Jacobs, Lamb, and Philipp (2010) are: (1) Please describe in detail what *you think* this child did in response to this problem, (2) Please explain what you learned about this child's understanding of mathematics, and (3) Pretend that you are the teacher of this child. What problems or questions might you pose next? Provide a rationale for your answer. In subsequent semesters, the italicized words were removed from the prompt to emphasize addressing the factual evidence of the video clip, not assumptions. The post-assessment task and protocol for delivery was identical to the pre-assessment. Administration of the post-assessment occurred within the last two weeks of the semester.

Construction of Noticing Benchmarks

The research team reviewed the professional noticing video segment and identified key response features for the attending prompt. We examined PSET attending responses from a single institution to identify emerging themes (Glaser & Strauss, 1967). Themes that emerged from this analysis included:

- Identifying key, salient activity (i.e., " . . . the child counted the bears, and then counted up to the amount of shells on his fingers")
- Identifying additional activity (i.e., " . . . he then looked to see how many fingers he had up")
- Operational presumptions (i.e., ". . . he subtracted 11-7")

Van Zoest, L. R., Lo, J.-J., & Kratky, J. L. (Eds.). (2012). *Proceedings of the 34th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education.* Kalamazoo, MI: Western Michigan University.

- Purporting evidence that did not occur in the segment (i.e., " . . . the child counted back from 11 to 7")
- Cognitive interpretations (i.e., ". . . the child lacks a sense of cardinality")

We created the themes based on the emerging patterns in order to co-construct a set of initial common benchmarks. The identification of emergent themes played a significant role during this initial drafting process as the characteristics of PSET attending responses coupled with researcher-identified key features suggested four distinct response types (elaborate, salient, limited, subordinate), ranging from a score of 4 to a score of 1. We scored our calibration data using these initial benchmarks. We replicated this process for the construction of the interpreting benchmarks and the deciding benchmarks. It is important to note that the coupling of emergent themes with researcher-identified key features resulted in three response categories, for the interpreting and deciding benchmarks, one fewer than the four natural levels that resulted for the attending benchmarks.

Scoring and Statistical Tests

Once the scoring benchmarks for each of the professional noticing questions were established, the research team scored all data. Each researcher scored one set (grouped by university) of data and a second researcher scored it again. Scores were compared and any discrepancies were discussed between the two researchers for a consensus. A third researcher was used in any case where a consensus was not reached.

An example of a time when a third researcher was needed consisted of a response in the attending question of the survey that stated the following: "I've learned that children, especially at younger ages, use their fingers a lot to help them count. It takes children a while to think mathematically in their head and not rely on objects to count." The first researcher felt this response deemed level 2 because the PSET recognized the progression of children's thinking from more concrete to more mental, however the second researcher felt it was too generic for this question and it should score a level 0 score. The second researcher felt if the response was specifically directed toward the student in the video, instead of children in general, then they would have agreed with the first researcher's score. When the third researcher scored this question, they also felt it was too generic and a score of 0 was given for this response.

Only participants with six scores consisting of the pre-assessment and post-assessment scores for each of the three components of professional noticing at the conclusion of the scoring process remained in the data set. Any participant who did not have six was removed, resulting in $n = 94$. *T*-test analyses were conducted to determine if significant changes occurred between the pre-assessment and post-assessment scores. This information was determined for the entire group as well as between the individual universities.

Results and Discussion

A one-way ANOVA was conducted to determine if there were overall statistically significant differences between the pre- and post- total scores for the professional noticing measure. Additionally, pre- and post-scores on each of the three individual professional noticing questions were tested for significance using a one-way ANOVA. The descriptive statistics, stratified by university as well as totals, are reported in Table 1 below. Also, Table 2 below shows the average gain in score from pre- to post-assessment, stratified by university as well as total participants.

Table 1: Descriptive Statistics of Professional Noticing Measures by University

			Attending		Interpreting		Deciding	
		N	*M*	*SD*	*M*	*SD*	*M*	*SD*
University A	Pre-Test	37	2.14	.79	1.59	.797	1.54	.61
	Post-Test	37	2.43	.87	2.05	.84	2.22	.79
University B	Pre-Test	23	2.39	.99	1.82	.89	2.04	.56
	Post-Test	23	3.09	1.04	2.43	.73	2.70	.56
University C	Pre-Test	34	2.38	1.10	1.76	.78	1.97	.67
	Post-Test	34	3.00	1.10	2.15	.89	2.47	.75
All Participants	Pre-Test	94	2.29	.96	1.71	.81	1.82	.66
	Post-Test	94	2.80	1.03	2.18	.84	2.43	.74

Table 2: Average Gains of Professional Noticing Measures by University

	Average Change			
	Attending	*Interpreting*	*Deciding*	*Total*
University A	.29	.46	.68	1.43
University B	.70	.61	.66	1.97
University C	.62	.39	.50	1.51
All Participants	.51	.47	.61	1.59

The means show growth on all three professional noticing measures between the pre-assessment and post-assessment measures at all universities. The increase was found to be significant ($F = 63.169$, $p < .001$) following tests to determine whether there is a statistically significant difference between the total pre- and post- scores of professional noticing. Interaction between the total scores and each university was not found to be significant ($F = .493$, $p = .612$). This is a positive result in that each university in the study is showing gains consistent with other universities.

A one-way ANOVA was conducted for each of the three questions to determine whether statistically significant gains were found for each component of professional noticing (attending, interpreting, deciding). The results of the ANOVA are found in Table 3 below. The smaller *F*-value found for the attending component can be attributed to the larger scale for that question (4-point scale) when compared to the interpreting and deciding questions (3-point scale).

Table 3: Results of ANOVA Comparing Pre- and Post-Assessments of All Universities

	Scale	*N*	*F*	*p*
Attending	1-4	94	61.43	< .001
Interpreting	1-3	94	1075.92	< .001
Deciding	1-3	94	1014.84	< .001

The interpreting component demonstrated the smallest change in growth overall ($\Delta = .47$), while the deciding component found the largest change in growth ($\Delta = .61$). Although significance was found between the pre- and post-scores for interpreting, this was the most difficult benchmark to construct due to the diversity of the responses, so the researchers are optimistic that a suitable benchmark has been created to adequately measure the PSETs interpreting skills. PSETs reported that the five-session module was repetitive in instruction for attending and interpreting and did not include enough instruction on the deciding component. They believed they needed further instructions on how to determine the next steps in mathematical questioning and instructional tasks to advance students into a higher SEAL stage. The five-session module was reduced to a four-session module (to reduce length and repetitiveness) and a stronger emphasis on instructional deciding was included. Future analyses with the new module and further refinement of the scoring benchmarks could provide more accurate assessments for growth in the three areas of professional noticing.

A statistically significant difference was not found for the attending component ($F = .519$, $p = .597$) and the deciding component ($F = 2.187$, $p = .118$) when comparing the three components of professional noticing within the three universities in the study. However, a statistically significant difference was found ($p < .05$) between the universities for the interpreting component ($F = 3.962$, $p = .022$). This lack of a statistically significant difference between the universities in attending and deciding is a positive result because it informs the researchers that the PSETs in the study are equally distributed in those components. In order to examine the statistically significant difference in the interpreting component, a Tukey's post-hoc analysis was conducted. This test suggested that University A's scores were the cause of the significance when compared within the overall data. This seems realistic considering University A's scores in the interpreting component were lower than University B and C. However, an independent means t-test was conducted for each of the professional noticing components in efforts to further investigate the scores from the individual universities.

The data was stratified by university and University A had statistically significant differences at $p < .05$ between pre- and post-assessments for interpreting ($t = -2.217$, $p = .033$) and at $p < .01$ for deciding ($t = -4.104$, $p = .000$). Attending was not statistically significant ($t = -1.571$, $p = .125$). University B had statistically significant differences at $p < .05$ for attending ($t = -2.729$, $p = .012$) and interpreting ($t = -2.440$, $p = .023$), and at $p < .01$ for deciding ($t = -4.035$, $p = .001$). University 2 had similar results with statistical significance at $p < .05$ for attending ($t = -2.670$, $p = .012$) and interpreting ($t = -2.196$, $p = .035$), and significance at $p < .01$ for deciding ($t = -3.253$, $p = .003$). Despite the fact that University A was statistically significantly lower overall than Universities B and C in interpreting based on the ANOVA test, the *t*-tests still reveal significance between pre- and post-assessments in interpreting for University A. The lack of significance in the attending component for University A is not surprising either, considering the PSETs from University A demonstrated the smallest amount of growth when compared to Universities B and C.

It should be noted that in cases where students scored high on the pre-assessment, they would not be expected to show growth. For example, in the attending component, 37% of students (12 of 37) from University A scored at least a 3 on the pre-assessment indicating they attended to the most salient features of the video assessment. University B and C had similar results with 39% (9 of 23) and 38% (13 of 34), respectively on the pre-assessment. In the interpreting component, 19% (7 of 37) from University A, 30% (7 of 23) from University B, and 21% (7 of 34) from University C all scored the highest possible score on the pre-assessment. The results in the deciding component were similar with 5% (2 of 37) from University A, 17% (4 of 23) from University B, and 21% (7 of 34) from University C all scoring perfect scores in the pre-assessment for instructional deciding, thus not allowing those students to show growth in that component. The researchers see the limited upper range as an opportunity to further refine the scoring benchmarks to include additional range, allowing for the sector of PSETs who scored perfect pre-assessment scores to show growth.

Final Remarks

In summary, preliminary findings at three sites suggest the efficacy of a researcher-developed module aimed at promoting professional noticing capacities among PSETs in the area of early number and operation. The development of such capacities among aspiring teachers at multiple sites bodes well for scaled establishment of responsive teaching practices within teacher preparation programs. Towards this end, subsequent module implementation and measurement will occur in teacher preparation programs at two additional institutions (for a total of five implementation sites); moreover, two different institutions have been identified to serve as non-implementing comparison sites.

References

Boerst, T. A., Sleep, L., Ball, D. L., & Bass, H. (2011). Preparing teachers to lead mathematics discussions. *Teachers College Record, 113*(12), 2844–2877.

Carnegie Foundation for the Advancement of Teaching. Retrieved from http://classifications.carnegiefoundation.org/lookup_listings/view_institution

Carpenter, T. P., Fennema, E., Franke, M. L., Levi, L. & Empson, S. B. (1999). *Children's mathematics: Cognitively guided instruction*. Portsmouth, NH: Heinemann.

Clements, D. H. (2007). Curriculum research: Toward a framework for "research-based curricula." *Journal for Research in Mathematics Education, 38*, 35–70.

Glaser, B., & Strauss, A. (1967). The discovery of the grounded theory: Strategies for qualitative research. New York: Aldine de Gruyter.

Grossman, P., Compton, C., Igra, D., Ronfeldt, M., Shahan, E., & Williamson, P. (2009). Teaching practice: A cross-professional perspective. *Teachers College Record, 111,* 2055–2100. Retrieved from http://www.tcrecord.org (ID Number 15018)

Jacobs, V. A., Lamb, L. L. C., & Philipp, R. A. (2010). Professional noticing of children's mathematical thinking. *Journal for Research in Mathematics Education, 41*(2*),* 169–202.

Kersting, N. B., Givvin, K. B., Sotelo, F. L., & Stigler, J. W. (2010). Teachers' analyses of classroom video predict student learning of mathematics: Further explorations of a novel measure of teacher knowledge. *Journal of Teacher Education, 61*, 172–181.

Mulligan, J., Bobis, J., & Francis, C. (1999). Insights into early numeracy: The Count Me in Too Project. *Australian Primary Mathematics Classroom, 4*, 22–27.

National Council of Teachers of Mathematics. (2006). *Curriculum focal points for prekindergarten through grade 8 mathematics*. Reston, VA: Author.

Schifter, D., Bastable, V., & Russell, S. J. (2000). *Developing mathematical ideas*. Columbus, OH: Pearson Education.

Seago, N., Mumme, J., & Branca, N. (2004). *Learning and teaching linear functions: Video cases for mathematics professional development*. Portsmouth, NH: Heinemann.

Sherin, M. G., Jacobs, V. R., Philipp, R. A. (Eds.). (2011). Mathematics teacher noticing: Seeing through teachers' eyes. New York: Routledge.

Sherin, M. G., & van Es, E. A. (2009). Effects of video club participation on teachers' professional vision. *Journal of Teacher Education, 60,* 20–37.

Star, J. R., & Strickland, S. K. (2008). Learning to observe: Using video to improve preservice mathematics teachers' ability to notice. *Journal of Mathematics Teacher Education, 11,* 107–125.

Steffe, L. (1992). Learning stages in the construction of the number sequence. In J. Bideaud, C. Meljac, & J. Fischer (Eds.), *Pathways to number: Children's developing numerical abilities* (pp. 83–88). Hillsdale, NJ: Lawrence Erlbaum.

Steffe, L. P., Cobb, P., & von Glasersfeld, E. (1988). *Construction of arithmetical meanings and strategies*. New York: Springer-Verlag.

Steffe, L. P., von Glasersfeld, E., Richards, J., & Cobb, P. (1983). *Children's counting types: Philosophy, theory, and application*. New York: Praeger Scientific.

Thomas, J. (2010). *Picture this: A doctoral dissertation examining children's quantitative mental imagery* (Unpublished doctoral dissertation). The University of Cincinnati.

Wright, R. J., Martland, J., & Stafford, A. (2006). *Early numeracy: Assessment for teaching and intervention* (2nd ed.). London, UK: Paul Chapman/Sage.

Van Zoest, L. R., Lo, J.-J., & Kratky, J. L. (Eds.). (2012). *Proceedings of the 34th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education.* Kalamazoo, MI: Western Michigan University.